Hybrid Methods Reveal Multiple Flexibly Linked DNA Polymerases Within the Bacteriophage T7 Replisome

Jamie R. Wallen[1,*], Hao Zhang[2], Caroline Weis[3], Weidong Cui[2], Brittni M. Foster[1], Chris M. W. Ho[4], Michal Hammel[3], John A. Tainer[5,3], Michael L. Gross[2], and Tom Ellenberger[4,*]

SUPPLEMENTAL INFORMATION

SUPPLEMENTAL DATA

donor	acceptor	occupancy
ARG590-Side	**PHE566-Main**	84.3%
SER546-Side	GLU495-Side	52.1%
THR293-Side*	ALA489-Main*	45.8%
ASN493-Side	**SER545-Main**	34.8%
SER281-Main*	GLU635-Side*	26.4%
LYS594-Main	**PHE566-Side**	26.1%
LYS589-Side	**ASP563-Side**	20.6%
ASP565-Main	THR572-Side	20.5%
ARG219-Side*	**GLU374-Side***	19.2%
SER546-Main*	GLU495-Side*	17.8%
ARG687-Side	**ASP297-Side**	17.4%
LYS377-Side	THR205-Side	14.0%
ARG490-Side	**GLU550-Side**	13.5%
SER281-Side*	GLU635-Side*	13.4%
THR293-Main	ASN493-Main	13.1%
ARG219-Side	**GLU371-Side**	12.8%
ARG370-Side	ASP682-Side	8.9%
ARG590-Side	**ASP563-Side**	7.0%
ARG490-Side	**ASP565-Side**	6.9%
ARG359-Side	GLU195-Side	6.7%
HIS498-Side	**GLU543-Side**	5.7%
ARG678-Side	**ALA332-Main**	5.4%
ARG590-Side	**ASN562-Main**	5.4%

Table SI. Related to Figure 3. Hydrogen-Bonding Interactions Between Pol B and gp4. A molecular dynamics simulation of Pol B interacting with gp4 was performed using GROMACS (see Methods). The most prevalent (highest occupancy) hydrogen-bonding interactions between Pol B and gp4 are listed here as the percentage of frames from the molecular dynamics trajectory (5000 frames total = 50 ns) showing the interaction. These hydrogen-bonding interactions involve either main chain or side chain atoms as noted. Gp4 residues participating in these interactions are shown in boldface, with residues from the tail of gp4 (residues 545-566) contributing many of the high occupancy hydrogen-bonding interactions with Pol B (highlighted in orange). Residues in Pol B and gp4 that make interactions not observed with Pol C (Table S2) are indicated with asterisks.

donor	acceptor	occupancy
SER365-Side	ASP682-Side	55.84%
ARG678-Side	**GLU371-Side**	54.68%
LYS628-Side*	**GLU538-Side***	28.25%
ASN493-Side	**GLU283-Side**	24.31%
LYS589-Side	**GLU543-Side**	16.09%
SER553-Side	GLU195-Side	12.06%
GLN571-Side	**GLU543-Side**	12.04%
LYS189-Side	**GLU550-Main**	9.44%
ARG359-Side	GLU567-Side	8.92%
GLN328-Side	ARG490-Main	8.34%
LYS189-Side*	**GLU551-Side***	7.20%
THR205-Side	**ASP509-Side**	5.98%
ARG687-Side	**ASP366-Side**	5.46%
ARG361-Side	ASP682-Side	5.32%

Table S2. Related to Figure 3. Hydrogen-Bonding Interactions Between Pol C and Gp4. Hydrogen-bonding interactions between Pol C and gp4 were identified with a molecular dynamics simulation using GROMACS (see Methods). High occupancy hydrogen bonds are listed here, based on the percentage of frames from the molecular dynamics trajectory (5000 frames total = 50 ns) showing the interaction. These hydrogen-bonding interactions involve either main chain or side chain atoms as noted. Gp4 residues participating in these interactions with Pol C are shown in boldface. The position of Pol C on the gp4 helicase precludes interactions with the C-terminal tail of gp4 (*cf*. Figure 2B), which are important for interactions with Pol A and Pol B (Figure 2A; Table SI). Although many of the same residues in Pol B and Pol C contact gp4, their interactions are different because of the rotated orientation of Pol C relative to Pol B (compare Figures 3A and 3B). Residues in Pol C and gp4 that make interactions not observed with Pol B (Table SI) are indicated with asterisks.

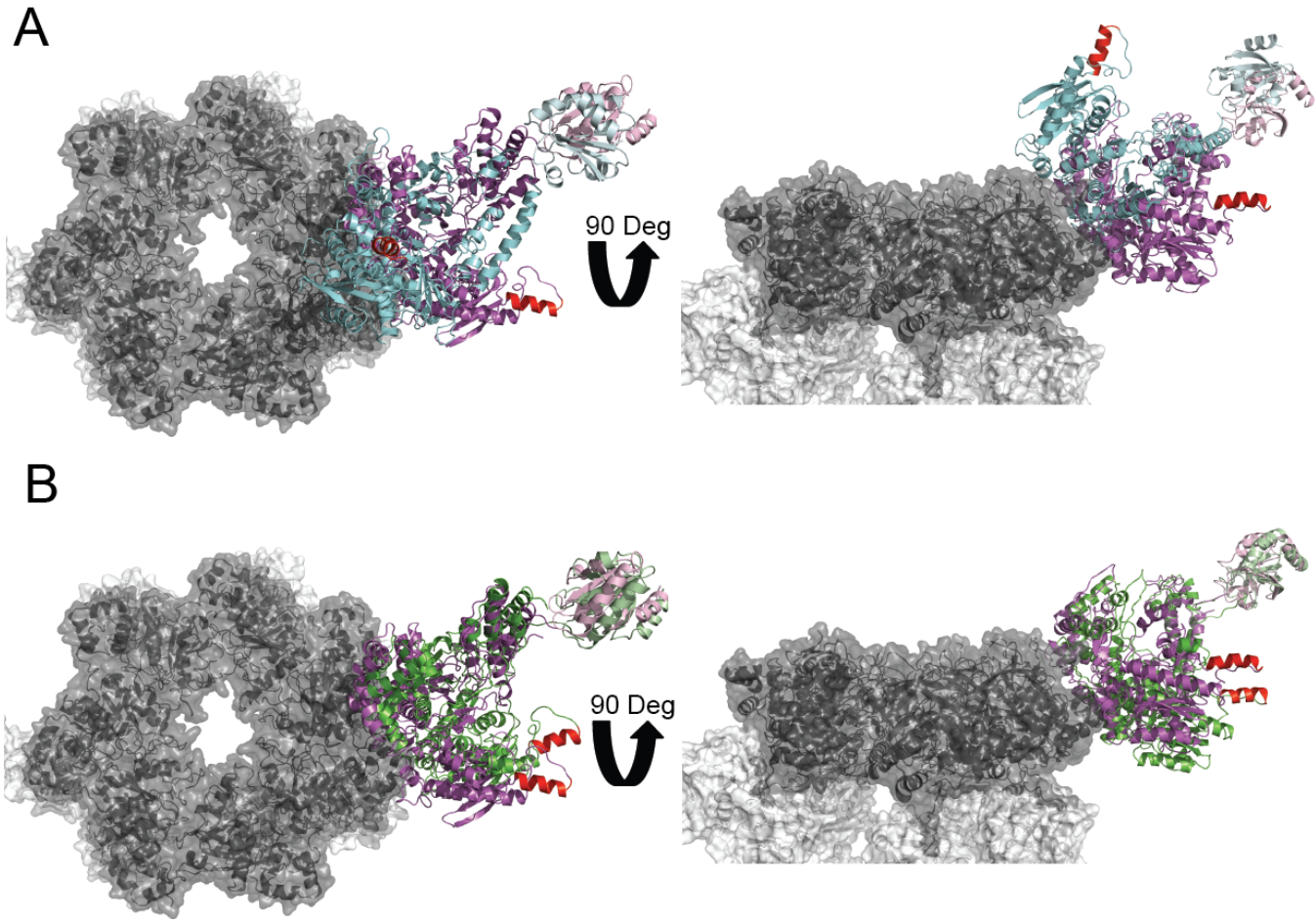


Figure S1. Related to Figures 1-3. Comparison of the Different Polymerase Binding Orientations on the Gp4 Heptameric Ring. (**A**) In the crystal structure, gp4 chains F and G interact with Pol B and Pol C, respectively, via the helix turn helix motif (residues 366-387) in the helicase domain that is highlighted in Figure 1. Pol C (cyan) was overlayed onto Pol B (magenta) by superimposing a complex of gp4 chain G-Pol C onto gp4 chain F (RMSD = 3.16 Å for the superposition of the gp4 chains) in the full structure. The binding orientation of Pol C differs from Pol B by an approximately 90-degree rotation around an axis that is tipped slightly out of the plane of the heptameric ring. A view from the C-terminal face of the heptameric ring as well as a 90-degree rotation shows the binding orientations of Pol B and Pol C in the structure. Residues 140-151, which form a small alpha helix in the exonuclease subdomain of the polymerase, are colored red in both Pol B and Pol C in order to highlight the different binding orientations. (**B**) A complex of gp4 chain D-Pol A was also superimposed on gp4 chain F (RMSD = 3.04 Å for the superposition of the gp4 chains) to show the similar binding orientations between Pol A and Pol B. Residues 140-151 of the polymerases are colored red. Although some movement is possible, both Pol A and Pol B bind gp4 in a similar manner and engage the same residues on the helicase ring.

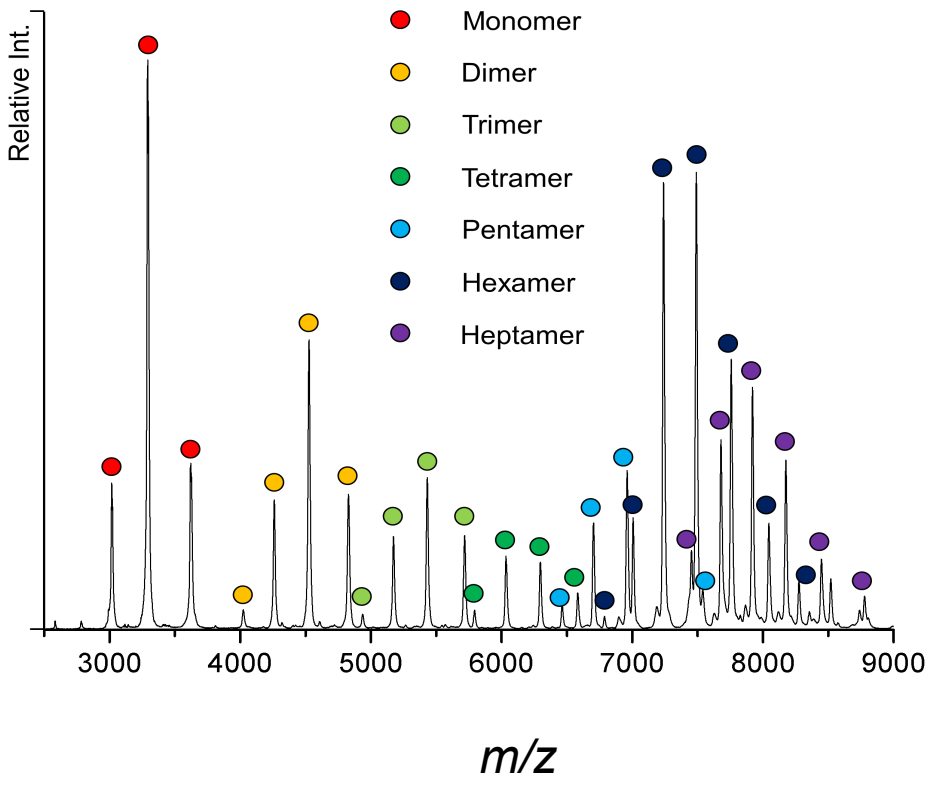
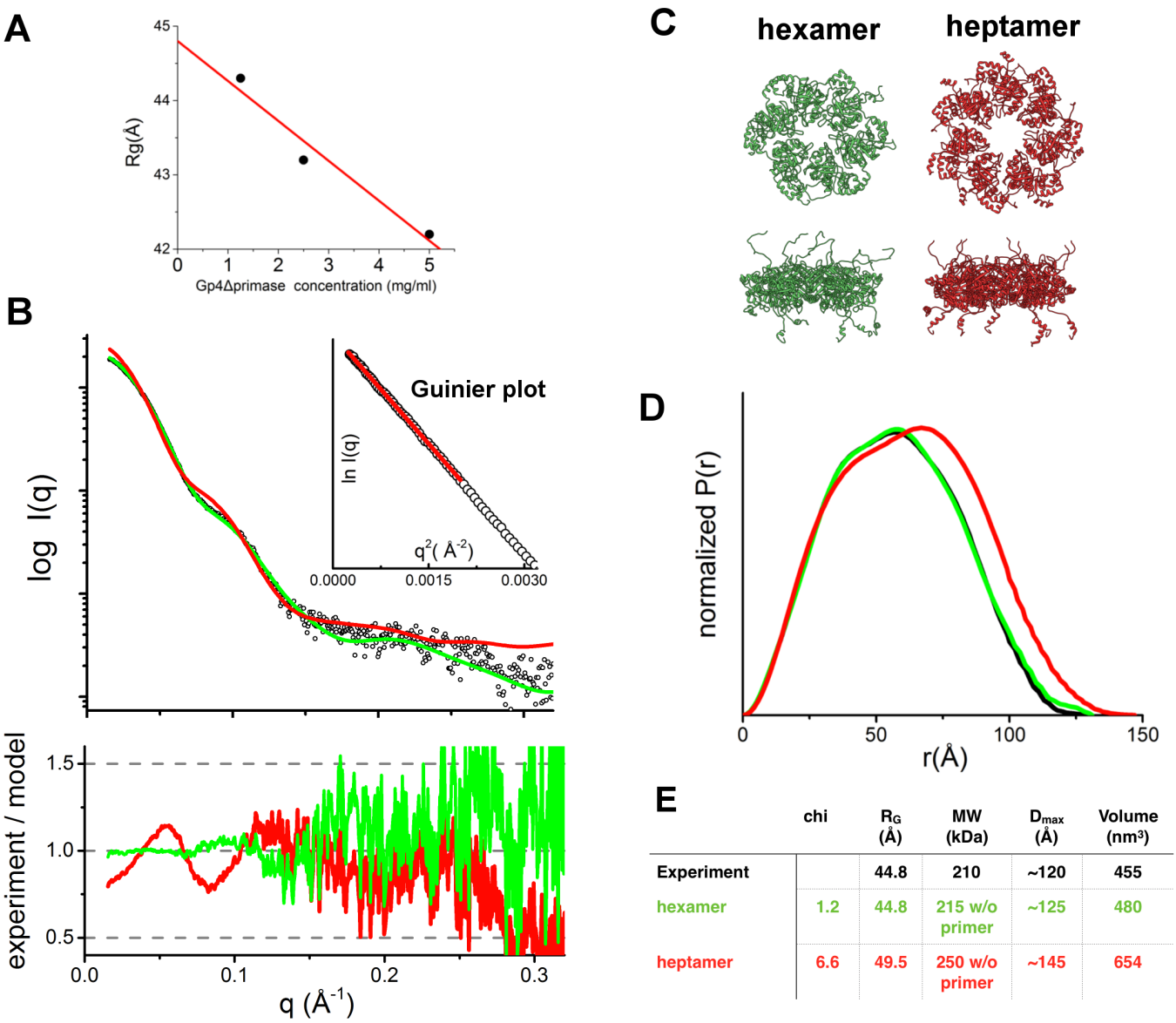


Figure S2. Related to Figure 4. Native Mass Spectrometry of Gp4Δprimase. The gp4Δprimase forms a variety of oligomeric species in solution. Peaks are color-coded according to oligomeric state, ranging from monomers to heptamers. Peak heights correspond to the relative abundance of each oligomeric species.



E

	chi	R_G (Å)	MW (kDa)	D_{max} (Å)	Volume (nm^3)
Experiment		44.8	210	~120	455
hexamer	1.2	44.8	215 w/o primer	~125	480
heptamer	6.6	49.5	250 w/o primer	~145	654

Figure S3. Related to Figure 5. Gp4Δprimase is Predominately a Hexamer in the Presence of DNA and Nucleotide Substrates. (**A-E**) SAXS analysis of gp4Δprimase alone shows that in the presence of a dT15 oligo and dTTP the gp4 protein preferentially assembles as a hexamer in solution. (**A**) Dependence of the radius of gyration (R_g) on gp4Δprimase concentration. The calculated R_g value for an infinitely dilute sample (R_g = 44.8 ± 0.4) matches the theoretical R_g value for the gp4Δprimase hexamer (see panel E). (**B**) A comparison of the experimental scattering profile of the gp4Δprimase (black) with a theoretical scattering profile obtained from the hexameric (green; χ = 1.2) and the heptameric (red; χ = 6.6) models of gp4. Inset: Guinier plot for experimental data at the limit of (q x Rg) < 1.6. (**C**). Two orthogonal views of atomic models for hexameric (1E0K, Singleton et al., 2000) and heptameric (1Q57, Toth et al., 2003) forms of gp4Δprimase. (**D**) The experimental P(r) of the gp4Δprimase (black) was compared with a theoretical P(r) obtained from models of the gp4 hexamer (green) and heptamer (red). (**E**) Experimentally determined SAXS parameters for the gp4Δprimase are compared to the calculated parameters for models of the hexamer and heptamer. The experimental data are in good agreement with values for the hexamer model.

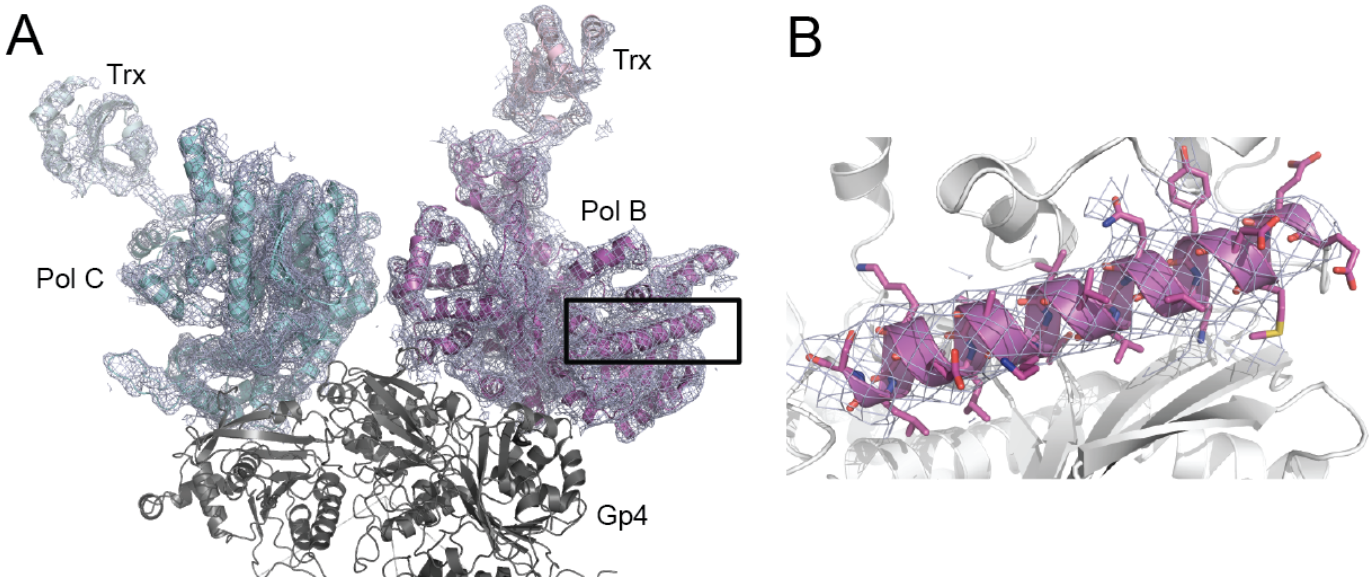


Figure S4. Related to Figures 1-3. Representative Electron Density Maps Showing the Placement of Pol B and Pol C in the Crystal Structure. (**A**) A $2F_o$-F_c map (blue) contoured at 1 sigma is shown for both Pol B (magenta) and Pol C (cyan). The three subunits of gp4 that contact Pol B and Pol C are colored gray. (**B**) A zoomed-in view of the boxed region of Pol B in panel (**A**) shows $2F_o$-F_c electron density contoured at 1 sigma for residues 165-187 of the exonuclease domain.

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Proteins

The exonuclease-deficient T7 DNA polymerase $gp5_{5A7A}$ (Patel et al., 1991) was used for all experiments and was purified as previously described (Doublié et al., 1998). A variant of the T7 primase-helicase protein lacking the zinc-binding domain of the primase (gp4ΔZBD, spanning residues 64-566) was used for crystallization studies, and it expressed at low levels most probably because of indiscriminate hydrolysis of nucleoside triphosphate pools in the *E. coli* host. To increase protein yield, the coding sequence for gp4ΔZBD was subcloned between the NcoI and HindIII sites of the pETDuet vector together with a gp4 deletion construct (gp4Δ271 spanning residues 272-566) (Guo et al., 1999) with impaired nucleotide hydrolysis activity, which was subcloned between the NdeI and XhoI sites of the same pETDuet vector. Co-expression of these gp4 proteins in BL21(DE3)pLysS cells substantially increased the yield of the gp4ΔZBD protein, most likely due to the formation of mixed oligomers of gp4ΔZBD and gp4Δ271 lacking nucleotide hydrolysis activity. Gp4ΔZBD was purified away from gp4Δ271 by 10% polyethylene glycol 4000 precipitation followed by loading onto a phosphocellulose column equilibrated in 20 mM potassium phosphate pH 6.8, 20 mM KCl, 1 mM EDTA, 10% glycerol, and 1 mM DTT. The gp4Δ271 does not bind to phosphocellulose and is removed during the phosphocellulose step. The column was washed and gp4ΔZBD was eluted using a 20-1000 mM KCl gradient. Fractions containing gp4ΔZBD were identified by SDS-PAGE and dialyzed into 20 mM Tris pH 7.5, 0.5 mM EDTA, 10% glycerol, and 0.5 mM DTT. The protein was then loaded onto a 10/100 MonoQ column equilibrated in 20 mM Tris pH 7.5, 100 mM NaCl, 0.5 mM EDTA, 10% glycerol, and 0.5 mM DTT. The column was washed and gp4ΔZBD was eluted using a 100-800 mM NaCl gradient. Fractions containing gp4ΔZBD were pooled, concentrated, and then loaded onto a Superdex 200 column equilibrated in 20 mM Tris pH 7.5, 200 mM NaCl, 0.5 mM EDTA, 10% glycerol, and 0.5 mM DTT.

The gp4Δprimase protein (residues 241 to 566) used for native MS and SAXS experiments was subcloned between the NdeI and XhoI sites of pET21b and expressed in BL21(DE3) cells. Gp4Δprimase was purified by 10% polyethylene glycol 4000 precipitation followed by loading onto a 5 mL HiTrap Q column equilibrated in 20 mM Tris pH 7.5, 100 mM NaCl, 0.5 mM EDTA, 10% glycerol, and 0.5 mM DTT. The column was washed and gp4Δprimase was eluted using a 100-1000 mM NaCl gradient. Fractions containing gp4Δprimase were identified by SDS-PAGE and were diluted to lower the NaCl concentration to 100 mM, and the protein was subsequently loaded onto a 10/100 MonoQ column equilibrated in 20 mM Tris pH 7.5, 100 mM NaCl, 0.5 mM EDTA, 10% glycerol, and 0.5 mM DTT. The column was washed and gp4Δprimase was eluted using a 100-800 mM NaCl gradient. Fractions containing gp4Δprimase were pooled, concentrated, and then loaded onto a Superdex 200 column equilibrated in 20 mM Tris pH 7.5, 200 mM NaCl, 0.5 mM EDTA, 10% glycerol, and 0.5 mM DTT. To generate a gp4Δprimase protein deficient in nucleotide hydrolysis, two point mutations (E343Q (Crampton et al., 2006a) and H465Y (Crampton et al., 2006b)) were introduced by site-directed mutagenesis. The gp4Δprimase (E343Q, H465Y) double mutant was expressed and purified as described for wild-type gp4Δprimase.

Structure Determination and Model Refinement

The structure of the T7 replisome was determined by molecular replacement using the program PHASER (McCoy et al., 2007), with coordinates for gp5/trx (PDB 2AJQ), the helicase domain of gp4 (PDB 1CR0) and the RNA polymerase domain of gp4 (PDB 1NUI) as starting models (Kato et al., 2003; Sawaya et al., 1999). PHASER correctly located the positions of all three gp5 subunits, six helicase domains, and six primase domains in the asymmetric unit. The thioredoxin subunit (trx) of T7 DNA polymerase, the seventh gp4 helicase domain, and the seventh primase domain were subsequently manually docked into the electron density calculated from the molecular replacement solution. Linkers connecting the primase and helicase domains were built using the gp4ΔZBD structure coordinates (PDB 1Q57) (Toth et al., 2003). Rigid body refinement of the crystallographic model was initially performed in REFMAC (Murshudov et al., 1997) before xyz coordinate and translation/libration/screw (TLS) model refinement in PHENIX (Adams et al., 2010). During model refinement, non-crystallographic symmetry (NCS) restraints were applied to all atoms within alpha-helices or beta-strands to maintain proper geometry while refining against the low-resolution data. The exonuclease (residues 1-199), thumb (residues 200-440), palm (residues 441-476, 645-704) and fingers (residues 477-644) subdomains of gp5 were treated as separate NCS groups. Each of the primase and helicase domains of gp4 were treated as separate NCS groups, and trx was treated as a NCS group separate from gp5. The TLS groups used in refinement were defined the same as the NCS restraints described above except that atoms making up flexible regions of the structure were also included in the TLS groups. In an effort to reduce phase bias resulting from the molecular replacement solution, the initial model of the T7 replisome was built into a series of omit electron density maps calculated by removing portions of the model during rigid body refinement then performing NCS averaging and solvent flattening using DM and Parrot (Cowtan, 2010; Cowtan and Main, 1996). The resulting omit maps revealed clear electron density for the omitted protein residues. Calculations of buried surface area were performed using the program PISA (Krissinel and Henrick, 2007). Representative electron density maps for Pol B and Pol C are provided in Figure S4.

Molecular Dynamics Simulations

All molecular dynamics simulations were run at 300 K using the AMBER ff03 force field (Kollman, 1996) with the TIP3P explicit solvent model (Jorgensen and Madura, 1983). The starting conformation used for this simulation was generated by placing the crystallographic structure in a dodecahedral water box that extended one nm beyond the protein in any dimension. Sodium counterions were added to achieve electrostatic neutrality. This system was energy minimized with the steepest descent algorithm until the maximum force fell below 1000 kJ/mol/min using a step size of 0.01 nm and a cutoff distance of 1.2 nm for the neighbor list, Coulomb interactions, and van der Waals interactions. The system was equilibrated for 1.0 nanosecond under periodic boundary conditions using the particle-mesh Ewald method for the long-range evaluation of electrostatic forces (Essmann et al., 1995). For production runs, all bonds were constrained with the LINCS algorithm (Hess, 2007) and virtual sites (Feenstra et al., 1999) were used to allow a 4 fs time step. Cut-offs of 1.0 nm were used for the neighbor list, Coulomb interactions, and van der Waals interactions. The Verlet cutoff scheme was used for the neighbor list. The stochastic velocity rescaling (v-rescale) thermostat (Bussi et al., 2007) was used to hold the temperature at 300 K. Conformations were stored every 10ps.

SUPPLEMENTAL REFERENCES

Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, et al., (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta. Crystallogr. D Biol. Crystallogr. *66*, 213–221.

Bussi, G., Donadio, and D., Parrinello, M. (2007). Canonical sampling through velocity rescaling. J. Chem. Phys. *126*, 014101.

Cowtan, K. (2010). Recent developments in classical density modification. Acta. Crystallogr. D Biol. Crystallogr. *66*, 470–478.

Cowtan, K.D., and Main, P. (1996). Phase combination and cross validation in iterated density-modification calculations. Acta. Crystallogr. D Biol. Crystallogr. *52*, 43–48.

Crampton, D.J., Mukherjee, S., and Richardson, C.C. (2006a). DNA-induced switch from independent to sequential dTTP hydrolysis in the bacteriophage T7 DNA helicase. Mol. Cell *21*, 165–174.

Crampton, D.J., Ohi, M., Qimron, U., Walz, T., and Richardson, C.C. (2006b). Oligomeric states of bacteriophage T7 gene 4 primase/helicase. J. Mol. Biol. *360*, 667–677.

Doublié, S., Tabor, S., Long, A.M., Richardson, C.C., and Ellenberger, T. (1998). Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 A resolution. Nature *391*, 251–258.

Essmann, U., Perera, L., Berkowitz, M.L., Darden, T., Lee, H., and Pedersen, L.G. (1995). A smooth particle mesh Ewald method. J. Chem. Phys. *103*, 8577–8593.

Feenstra, K.A., Hess, B., and Berendsen, H. (1999). Improving efficiency of large timescale molecular dynamics simulations of hydrogen-rich systems. J. Comput. Chem. *20*, 786-798.

Guo, S., Tabor, S., and Richardson, C.C. (1999). The linker region between the helicase and primase domains of the bacteriophage T7 gene 4 protein is critical for hexamer formation. J. Biol. Chem. *274*, 30303–30309.

Hess, B. (2007). P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. J. Chem. Theory and Comput. *4*, 116-122..

Jorgensen, W.L., and Madura, J.D., 1983. Quantum and statistical mechanical studies of liquids. Solvation and conformation of methanol in water. J. Am. Chem. Soc. *105*, 1407–1413.

Kato, M., Ito, T., Wagner, G., Richardson, C.C., and Ellenberger, T. (2003). Modular architecture of the bacteriophage T7 primase couples RNA primer synthesis to DNA synthesis. Mol. Cell *11*, 1349–1360.

Kollman, P.A. (1996). Advances and Continuing Challenges in Achieving Realistic and Predictive Simulations of the Properties of Organic and Biological Molecules. Acc. Chem. Res. *29*, 461–469.

E. Krissinel and K. Henrick (2007). Inference of macromolecular assemblies from crystalline state. J. Mol. Biol. *372*, 774--797.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J. Appl. Crystallogr. *40*, 658–674.

Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. Acta. Crystallogr. D Biol. Crystallogr. *53*, 240–255.

Patel, S.S., Wong, I., and Johnson, K.A. (1991). Pre-steady-state kinetic analysis of processive DNA replication including complete characterization of an exonuclease-deficient mutant. Biochemistry *30*, 511–525.

Sawaya, M.R., Guo, S., Tabor, S., Richardson, C.C., and Ellenberger, T. (1999). Crystal structure of the helicase domain from the replicative helicase-primase of bacteriophage T7. Cell *99*, 167–177.

Singleton, M.R., Sawaya, M.R., Ellenberger, T., and Wigley, D.B. (2000). Crystal structure of T7 gene 4 ring helicase indicates a mechanism for sequential hydrolysis of nucleotides. Cell *101*, 589–600.

Toth, E.A., Li, Y., Sawaya, M.R., Cheng, Y., and Ellenberger, T. (2003). The crystal structure of the bifunctional primase-helicase of bacteriophage T7. Mol. Cell *12*, 1113–1123.